To preprocess the two datasets, each country code of 'life.csv' was compared with the country codes of the other data set, and a new dataframe 'data' was created, consisting of the data of 'life expectancy' added to the data of 'world.csv'. Then, 'data' underwent median imputations, replacing the empty data points and finally standardized using StandardScaler(). From here, the accuracies of the algorithms were processed.

From the results of the accuracies of the three algorithms (decision trees and k-nn with k =5 and k =10), the decision tree performed better on this dataset than the k-nn algorithms. This may be because the target data 'life expectancy' is nominal data of 'high', 'medium' and 'low' which is more friendly to the algorithm that classifies data. On the other hand, k-nn is an algorithm which groups data together, but in this case, it may not necessarily relate to the target data 'life expectancy' directly.

K-nn at k=10 performed better than k-nn at k =5 (0.567>0.522), which may be due to the mere size of the k value. For a data set such as this, a larger k-value than 5 may be preferred over the smaller k-values as this allows for larger similar groups to be obtained, hence increasing the accuracy of the algorithm. However, this isn't necessarily a positive linear relationship with the accuracy of the algorithm. This is shown by testing the system at k = 20, the accuracy peaks at 0.6 but then decreases to 0.556 at k = 21.

For 2b, the same code from part 2a was utilized to preprocess the data to undergo the feature generating algorithms. In order to generate the interaction term pair features, every column in the processed dataframe 'data' was multiplied to each other, leaving out the pairs that already have been multiplied together and the pairs that are two of the same (hence performing a combination of 20 choose 2). This generated 190 features, which were all added to a new dataframe 'data'. To generate a feature by clustering, 3 clusters of all the data in the original csv file was used to make a column of value 0,1,2. Three clusters were chosen as it corresponds to the number of variables in life expectancy (high, medium and low). This was because the cluster was expected classify the datasets into clusters corresponding to the life expectancy variables. Finally, to select 4 features, every feature's chi square value was calculated with the cluster feature data which was used as the expected data and the 4 features who are not independent with the cluster feature and had the highest chi square value was used to calculate the accuracy.

However, the coding for 2b may have been severely faulty as the three results produced were highly unlikely to be true. This can be inducted from the fact that the accuracy of the first four features were the highest (0.633), the accuracy with the PCA algorithm was second (0.578) and the accuracy of feature engineering was the lowest (0.522). There may be a very slight chance that the accuracy of the first four features exceed the other processed accuracies, but this is highly unlikely. Also, the fact that the accuracy of the feature engineering is equivalent to the accuracy of the original data at k-nn at k=5 implies that this code is definitely faulty as the processed data should intuitively be more accurate.

A technique that could be implemented for this data would be utilize the greedy wrapper approach, where you compute the accuracy of every type of feature and choose the best group. However, this way is harsh in computational time and memory, hence may be unfit for a large data set.

From the results acquired, the reliance of the classification model cannot be judged as the code of the algorithm is not correctly implemented.